

5-9-2019

Analyzing Images, Editing Texts: The Manchester Project

Naima Afif

The University of Manchester, naima.afif@manchester.ac.uk

Siam Bhayro

University of Exeter, s.bhayro@exeter.ac.uk

Peter E. Poormann

The University of Manchester, peter.pormann@manchester.ac.uk

William I. Sellers

The University of Manchester, william.sellers@manchester.ac.uk

Natalia Smelova

The University of Manchester, natalia.smelova@manchester.ac.uk

Analyzing Images, Editing Texts: The Manchester Project

Abstract

This article discusses the methodologies and tools employed in the study of the Syriac Galen Palimpsest. While it focusses on the efforts of the ongoing Manchester Project, attention is also paid to earlier and contemporary work, particularly the most recent phase of research (which can be said to have started in 2009). In this way, the Manchester Project is properly contextualised. We describe the image analysis techniques employed by the Manchester team. The challenge is to reduce the information contained in the set of multi-spectral images and enhance it where it can usefully distinguish between undertext and overtext. One can either use unsupervised or supervised dimensional reduction techniques. An unsupervised method such as principle component analysis (PCA) provides an automatic result, whereas a supervised method such as Canonical Variates Analysis (CVA) requires one to teach the system by identifying blank areas, areas with only overtext, areas with only undertext, and areas with both. Using the resulting improvements to the visibility of the undertext, the Manchester team has been able to make significant advances in identifying where its folios fit into Galen's Book of Simple Drugs. The use of a program called SketchEngine is outlined, which permits an engagement with parallel Greek and Syriac texts and powerful searches - this is particularly useful for those folios that come from Books 6–8, for which a parallel Syriac manuscript exists. Having completed this initial stage, it became clear that around 100 folios that did not come from Books 6-8 remained to be identified. SketchEngine again has proved to be very useful in facilitating identifications of these folios. To illustrate the different challenges posed by these two distinct scenarios, examples are provided from Books 5 and 8.

Keywords

Galen, Syriac, Palimpsest, Sergius of Rēš ʿAynā, digital humanities, corpus linguistics, Greek

MANUSCRIPT STUDIES

A Journal of the Schoenberg Institute for Manuscript Studies

VOLUME 3, NUMBER 1
(*Spring 2018*)

Manuscript Studies (ISSN 2381-5329) is published semiannually
by the University of Pennsylvania Press



The Schoenberg Institute
for Manuscript Studies

UNIVERSITY of PENNSYLVANIA LIBRARIES

MANUSCRIPT STUDIES

A Journal of the Schoenberg Institute for Manuscript Studies

VOLUME 3, NUMBER 1

Special Issue:
The Syriac Galen Palimpsest
Editors: William Noel and Ralph M. Rosen

Articles

- The Syriac Galen Palimpsest Project: An Introduction
WILLIAM NOEL AND RALPH M. ROSEN 1
- Pulling It All Together: Managing the Syriac Galen
Palimpsest Project
MICHAEL B. TOTH 9
- The Codicology and Conservation of the Syriac
Galen Palimpsest
ABIGAIL B. QUANDT AND RENÉE C. WOLCOTT 33
- Spectral Imaging Methods Applied to the Syriac
Galen Palimpsest
ROGER L. EASTON, JR., KEITH T. KNOX, WILLIAM A. CHRISTENS-BARRY,
AND KEN BOYDSTON 69
- The Galen Palimpsest and the Modest Ambitions of the
Digital Data Set
DOUG EMERY 83
- The Syriac Galen Palimpsest: A Tale of Two Texts
NAIMA AFIF, SIAM BHAYRO, GRIGORY KESSEL, PETER E. PORMANN,
WILLIAM I. SELLERS, AND NATALIA SMELOVA 110

Analyzing Images, Editing Texts: The Manchester Project NAIMA AFIF, SIAM BHAYRO, PETER E. PORMANN, WILLIAM I. SELLERS, AND NATALIA SMELOVA	155
--	-----

The Textual Interest of the Syriac Versions of Galen's <i>Simples</i> IRENE CALÀ, JIMMY DACCACHE, AND ROBERT HAWLEY	186
--	-----

Annotations

Of Scribes and Scripts: Citizen Science and the Cairo Geniza LAURA NEWMAN ECKSTEIN	208
---	-----

Preserving Endangered Archives in Jerba, Tunisia: The al-Bāsi Family Library Pilot Project ALI BOUJDIDI AND PAUL M. LOVE	215
--	-----

The Intricacies of Capturing the Holdings of a Mosque Library in Yemen: The Library of the Shrine of Imām al- Hādī, Ṣa'da SABINE SCHMIDTKE	220
---	-----

Compilation, Collation and Correction in the Time of Encyclopedism: The Case of UPenn LJS 55 NATHALIE LACARRIÈRE	238
--	-----

Mapping Manuscript Migrations: Digging into Data for the History and Provenance of Medieval and Renaissance Manuscripts TOBY BURROWS, EERO HYVÖNEN, LYNN RANSOM, AND HANNO WIJSMAN	249
---	-----

Reviews

Aḥmad 'Abd al-Bāsiṭ. <i>Catalogue of the Private Collections of Manuscripts in the Egyptian National Library</i> ELIAS G. SABA	253
---	-----

David T. Gura. *A Descriptive Catalogue of the Medieval and Renaissance Manuscripts of the University of Notre Dame and Saint Mary's College*

LISA FAGIN DAVIS 256

Christopher De Hamel. *Meetings with Remarkable Manuscripts*

DANIEL TRAISTER 260

Analyzing Images, Editing Texts

The Manchester Project

NAIMA AFIF

The University of Manchester

SIAM BHAYRO

The University of Exeter

PETER E. PORMANN

The University of Manchester

WILLIAM I. SELLERS

The University of Manchester

NATALIA SMELOVA

The University of Manchester

THE SYRIAC GALEN PALIMPSEST (SGP) is currently the subject of a major project, funded by the Arts and Humanities Research Council (AHRC) and entitled “The Syriac Galen Palimpsest: Galen’s *On Simple Drugs* and the Recovery of Lost Texts through Sophisticated Imaging Techniques”;¹ it is based in Manchester, and the authors of the present article form the research team as it is currently constituted. In this article,

1 AH/M005704/1; September 2015 to February 2020.

we shall briefly tell the story of the manuscript and the various activities that have led up to the current project.² We shall then outline our methodologies for image analysis and philological study. In both areas, we use new digital tools to decipher the erased undertext. By combining methodologies from different disciplines, we are thus able to reconstruct the undertext. It is therefore no surprise that the project resulted from an intense collaboration between scholars, librarians, conservators, and imaging specialists, and we continue in this vein, as the present special issue amply illustrates. Let us begin by narrating the events that have led up to the current project.

Pre-history of the Project

Before the most recent stage of work on the SGP, the only known surviving witness to the Syriac version of Galen's *Book of Simple Drugs* was the manuscript London, British Library (BL) Add. 14661 (henceforth MS BL). It contains Sergius of Rēš 'Aynā's (d. 536) Syriac version of the Greek source text, but only for three out of the work's eleven books, namely Books 6 to 8.³ Each book is preceded by an introduction produced by Sergius, which included a dedication to his colleague Theodore, and an alphabetical list of

2 See also S. Bhayro and S. Brock, "The Syriac Galen Palimpsest and the Role of Syriac in the Transmission of Greek Medicine in the Orient," in *Ancient Medical and Healing Systems: Their Legacy to Western Medicine*, ed. R. David ([Manchester]: University of Manchester Library, [2013]), 25–43; and S. Bhayro, R. Hawley, G. Kessel, and P. E. Pormann, "The Syriac Galen Palimpsest: Progress, Prospects and Problems," *Journal of Semitic Studies* 58 (2013): 131–48.

3 BL Add. 14661 was the subject of another major project, funded by the European Research Council, that was based in Paris under the auspices of Robert Hawley—Floriental: From Babylon to Baghdad: Towards a History of the Herbal in the Near East (ERC-2010–StG-263783; September 2011 to August 2017). For how the AHRC and ERC projects complement each other, see N. Afif, C. Arsene, S. Bhayro, I. Calà, J. Daccache, R. Hawley, G. Kessel, P. E. Pormann, W. I. Sellers, and N. Smelova, "Continuing Research on the Syriac Galen Palimpsest: Collaborative Implementation Within the Framework of Two European Projects," *Semitica et Classica* 9 (2016): 261–68.

Greek plant names transliterated into Syriac.⁴ This last feature is distinctive and played an important part in our recent work on the SGP.

William Wright cataloged the manuscript and dated it to the sixth or seventh century, so it is not too far removed from Sergius himself.⁵ In 1885, Merx published a partial edition, which included Sergius's introductions accompanied by a German translation.⁶ Nearly a century later, Rainer Degen referred to the manuscript in two articles surveying the Syriac medical literature with special focus on Galen.⁷ In 2002, Siam Bhayro presented a conference paper on the importance of this manuscript for our study of Syriac medical terminology.⁸ Yet Gignoux did not incorporate it into the list of sources used in his 2011 lexicon of Syriac pharmacological terms.⁹ At that time, Peter E. Pormann explored the importance of MS BL for the study of Greek-Syriac-Arabic translation technique in two articles.¹⁰ Robert

4 These features are discussed in S. Bhayro, "Syriac Medical Terminology: Sergius and Galen's Pharmacopia," *Aramaic Studies* 3 (2005): 147–65 at 162–65; see also Robert Hawley's remarks in S. Bhayro and R. Hawley, "La littérature botanique et pharmaceutique en langue syriaque," in *Les sciences en syriaque*, ed. É. Villey (Paris: Geuthner, 2014), 285–318 at 296–98.

5 W. Wright, *Catalogue of the Syriac Manuscripts in the British Museum, acquired since the year 1838* (London: British Museum, 1872), 3:1187. See also G. Kessel, "Inventory of Galen's Extant Works in Syriac," in J. C. Lamoreaux, *Hunayn ibn Ishāq on His Galen Translations* (Provo: Brigham Young University, 2016), 168–92 at 176. The most up-to-date study of the manuscript is the article by N. Afif, S. Bhayro, G. Kessel, P. E. Pormann, W. I. Sellers, and N. Smelova, "The Syriac Galen Palimpsest: A Tale of Two Texts," contained in this special issue.

6 A. Merx, "Proben der syrischen Uebersetzung von Galenus' Schrift über die einfachen Heilmittel," *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 39 (1885): 237–305.

7 R. Degen, "Ein Corpus Medicorum Syriacorum," *Medizinhistorisches Journal* 7 (1972): 114–22 at 116, 120; R. Degen, "Galen im Syrischen: Eine Übersicht über die syrische Überlieferung der Werke Galens," in *Galen: Problems and Prospects. A Collection of Papers submitted at the 1979 Cambridge Conference*, ed. V. Nutton (London: The Wellcome Institute for the History of Medicine, 1981), 131–66 at 146–47.

8 Bhayro, "Syriac Medical Terminology."

9 See P. Gignoux, *Lexique des termes de la pharmacopée syriaque* (Paris: Association pour l'avancement des études iraniennes, 2011), 9–13.

10 P. E. Pormann, "The Development of Translation Techniques from Greek into Syriac and Arabic: The Case of Galen's *On the Faculties and Powers of Simple Drugs*, Book Six," in *Medieval Arabic Thought: Essays in Honour of Fritz Zimmermann*, ed. R. Hansberger, M. Afifi al-Akiti, and C. Burnett (London: Warburg Institute; Turin: Nino Aragno, 2012), 143–62;

Hawley and Jimmy Daccache then transcribed MS BL and kindly shared their transcription with us; this allowed us to identify leaves from Books 6 to 8 with relative ease, and we shall explain the modalities below.

Whereas MS BL contains only Books 6 to 8, the SGP has significantly more text. The Leipzig book dealer Karl W. Hiersemann offered it for sale in 1922, and the eminent German orientalist Anton Baumstark provided a first description in the auction catalog.¹¹ The undertext was clearly visible on certain pages, and Baumstark did not fail to arouse commercial interest by offering this hyperbolic description:

The No. 3 (partly), 20 [i.e., SGP] and 42 are palimpsests and we have reason to believe that the older writing . . . may be made readable. The realization of this possibility would be in case of No. 20 [sc. SGP] of the greatest value to the history of the medicine of the Syrians. . . . None of these treasures have as yet been utilized by science.

The manuscript was bought by the Swiss industrialist Arnold Mettler-Specker (1867–45) and after his death sold at auction in New York in 1948. It came into the possession of the current owner, who invested significantly in its conservation and imaging at the Walters Art Museum in Baltimore, Maryland, in 2009–10. Sebastian Brock identified a passage from Book 6 and thus established for the first time that the SGP contained Sergius's translation of Galen's *On Simple Drugs*. This in itself was a hugely significant discovery, and proved to be a watershed moment. Siam Bhayro later confirmed Brock's identification. This still only gave us an indication that the SGP probably contained Books 6 to 8, those also present in MS BL. But already in 2010, Bhayro surmised that the SGP probably originally contained all eleven books, and that it represents the most important Syriac

Peter E. Pormann, "The Formation of the Arabic Pharmacology: Between Tradition and Innovation," *Annals of Science* 68 (2011): 493–515.

11 K. W. Hiersemann, *Katalog 500: Orientalische Manuskripte* (Leipzig: Hiersemann, 1922), no. 20.

medical text so far discovered.¹² As we shall see, this suggestion was subsequently confirmed.

During this early phase, the materiality of the manuscript for the purpose of identification came into sharp focus. Two conservators, Abigail Quandt and Linda Owen, noticed that some of the leaves have section numbers in the margins, written in red ink and just visible to the naked eye; they subsequently made nearly forty identifications of such section numbers. At the end of the work, Bhayro proposed to call the manuscript the “Syriac Galen Palimpsest,” as he was confident that the whole manuscript contained Galenic material in Syriac translation. His suggestion was readily accepted and is now the common name.

From 2010, the huge array of digital images of the SGP was therefore in the public domain.¹³ This led to two parallel initiatives. William Noel asked Michael Sharp, the classics commissioning editor at Cambridge University Press, who had published the Archimedes Palimpsest, to suggest someone who could promote research on the SGP. Sharp put him in touch with Pormann, who then organized an international workshop on the SGP and also talked with Bill Sellers at the John Rylands Heritage Imaging Seminar in November 2013, resulting in a number of publications.¹⁴ Robert Hawley successfully applied to the European Research Council (ERC) in late 2009 for his project on Near Eastern herbals called *Floriental*; in it he already mentioned the aim of producing a critical edition of MS BL and the SGP (of which he learned from Bhayro). He organized a number of workshops and meetings, and identified various leaves from Books 6 to 8.¹⁵ Grigory

12 The reasons for this are discussed more fully in Bhayro and Brock, “The Syriac Galen Palimpsest.”

13 “The Galen Syriac Palimpsest,” www.digitalgalen.net, accessed 25 May 2017. The site is now hosted by OPenn, the University of Pennsylvania Libraries’ open-source repository (<http://openn.library.upenn.edu>).

14 S. Bhayro, R. Hawley, G. Kessel, and P. E. Pormann, “Collaborative Research on the Digital Syriac Galen Palimpsest,” *Semitica et Classica* 5 (2012): 261–64; S. Bhayro, P. E. Pormann, and W. I. Sellers, “Imaging the Syriac Galen Palimpsest: Preliminary Analysis and Future Prospects,” *Semitica et Classica* 6 (2013): 299–302.

15 R. Hawley, “More Identifications of the Syriac Galen Palimpsest,” *Semitica et Classica* 7 (2014): 237–72.

Kessel subsequently identified a few other leaves, including those from Books 2 and 4, and also found missing folios that originally belonged to the SGP in Harvard, Paris, Vatican, and St. Catherine's Monastery on Mount Sinai.¹⁶

This is a remarkable demonstration of the significance of the SGP: it contains text that is not otherwise preserved in Syriac, and it also potentially allows us to rewrite the history books. Sebastian Brock has described, on a number of occasions, the three distinct phases of Greek into Syriac translation activity—namely, those of the sixth century, the seventh century, and the late eighth to ninth centuries, and how the first and third phases were very different, partly due to the considerable lexicographical development that occurred over the three centuries that separated them.¹⁷ Ḥunayn, in his *Epistle*, stated that the first part (that is, Books 1 to 5) of Galen's *Book of Simple Drugs* was only translated into Syriac by Joseph the Priest (dates unknown); Job of Edessa (d. ca. 835); and finally Ḥunayn himself.¹⁸ Bhayro and Brock, however, have argued that because of the content of the preface to Book 6 by Sergius, he must also have produced a Syriac version of Books 1 to 5, not just 6 to 11, as Ḥunayn claimed.¹⁹ This raises the question of whether the translation of Books 2 to 5 that we have found in the SGP is by Sergius or not.

Given the nature of the material contained in Books 1 to 5, which is more discursive and argumentative, we now have an opportunity to assess in depth the dominant narrative of the relative value of the sixth-century translations. This process has already started, and some of us put it thus: "Ḥunayn is far more indebted to Sergius' efforts than one would guess

16 These two passages are equivalent to K. G. Kühn, ed., *Claudii Galeni, Opera omnia* (Leipzig: Car. Cnoblochii, 1821), 11:510, 658–59; see G. Kessel, "Membra disjecta sinaitica I: A Reconstitution of the Syriac Galen Palimpsest," in *Manuscripta Graeca et Orientalia: Mélanges monastiques et patristiques en l'honneur de Paul Géhin*, ed. A. Binggeli, A. Boud'hors, and M. Cassin (Leuven: Peeters, 2016), 469–96 at 489.

17 See, for example, S. P. Brock, *A Brief Outline of Syriac Literature* (Kottayam: St. Ephrem Ecumenical Research Institute, 1997), 116–19; S. Brock, "The Syriac Background to Ḥunayn's Translation Techniques," *ARAM* 3 (1991): 139–62 at 151.

18 See J. C. Lamoreaux, *Ḥunayn ibn Ishāq on His Galen Translations* (Provo, UT: Brigham Young University Press, 2016), 66–69.

19 Bhayro and Brock, "The Syriac Galen Palimpsest," 40.

from Ḥunayn's own account of how he rendered Galen into Syriac and Arabic."²⁰ To study the undertext of the SGP, we first needed to make the digital images produced in 2009–10 more readable, and here we enlisted the help of Bill Sellers.

Image Analysis

The goal of the technological support for the Syriac Galen Palimpsest project is primarily to enable the scholars to read the digital images of the undertext. Achieving this requires an efficient and repeatable methodology to produce readable text from the raw images. The image data available at www.digitalgalen.net has already been processed such that the undertext is often visible. The available images, however, have been reduced from the original 16-bit to 8-bit form. The process used to produce the 8-bit images was automated and did not allow manual optimization of the visible images. Therefore, we chose to work with the flattened 16-bit images (these have been corrected for uneven lighting by using standard gray flats illuminated using the same lighting and camera arrangement) before conversion to 8-bit. This gives us much more scope to customize the enhancement algorithms for specific images depending on what is needed by the scholar for a specific leaf, or region of a leaf.

The first step in any image enhancement process is to investigate what information the images contain. To illustrate this process, we used a 1000 × 1000 section of the 214v–221r page, and this section is used in all the subsequent discussion as representative of the manuscript as a whole. The data consists of thirteen images taken with different colors of LED illumination, four images taken with raking illumination, three images taken with UV illumination and color filters, and three images taken with blue illumination and color filters. The raking illumination images, however, are primarily useful in situations where the topography of the paper contains useful information where it shows up as shadows, and in the case of the

20 See Bhayro, Hawley, et al., "The Syriac Galen Palimpsest," 143.

SGP this does not seem to be the case. This leaves us with nineteen different images of the same page. These images are all accurately registered, and each image contains a different subset of the total information recorded from the page during the digitizing process. We can illustrate the information contained by manually selecting points on the image that correspond to specific parts of the content. Figure 1 shows the mean intensities recorded for the overtext, the undertext, and the parchment for one hundred points in the nineteen flattened 16-bit images. There are clear differences in the mean intensities, and this is of course why the overtext and undertext are visible on this image. The intensities recorded, however, vary depending on the frequency of light used to illuminate the page and the presence or absence of filters. In particular, histogram B makes it very clear that the biggest differences between the undertext and the parchment occur at low wavelengths, and this is why the undertext is much clearer in image C than in image D or E.

The simplest way of trying to improve the legibility of the undertext is to choose the images where we have the biggest differences between the intensities of the different components. Looked at individually, these can be translated into grayscale images, but we can also choose three images and apply them to the red, green, and blue channels of an image and thereby generate a pseudocolor image that can allow the eye to distinguish further between the different parts of the image.²¹ This process is illustrated in figure 2, and these images can be further enhanced in terms of brightness and contrast using standard photographic image software such as Adobe Photoshop (www.adobe.com) or more specialized scientific imaging software such as ImageJ (imagej.nih.gov). In many cases where multispectral imaging is used to visualize hidden features on manuscripts, this is all that is necessary to make the text readable, and it is certainly the first step that should be used when assessing what might be present.

21 R. L. Easton, K. T. Knox, and W. A. Christens-Barry, "Multispectral Imaging of the Archimedes Palimpsest," in *32nd Applied Imagery Pattern Recognition Workshop, Proceedings* (Los Alamitos, CA: IEEE Computer Society, 2003), 111–16; doi: 10.1109/AIPR.2003.1284258, accessed 25 May 2017.

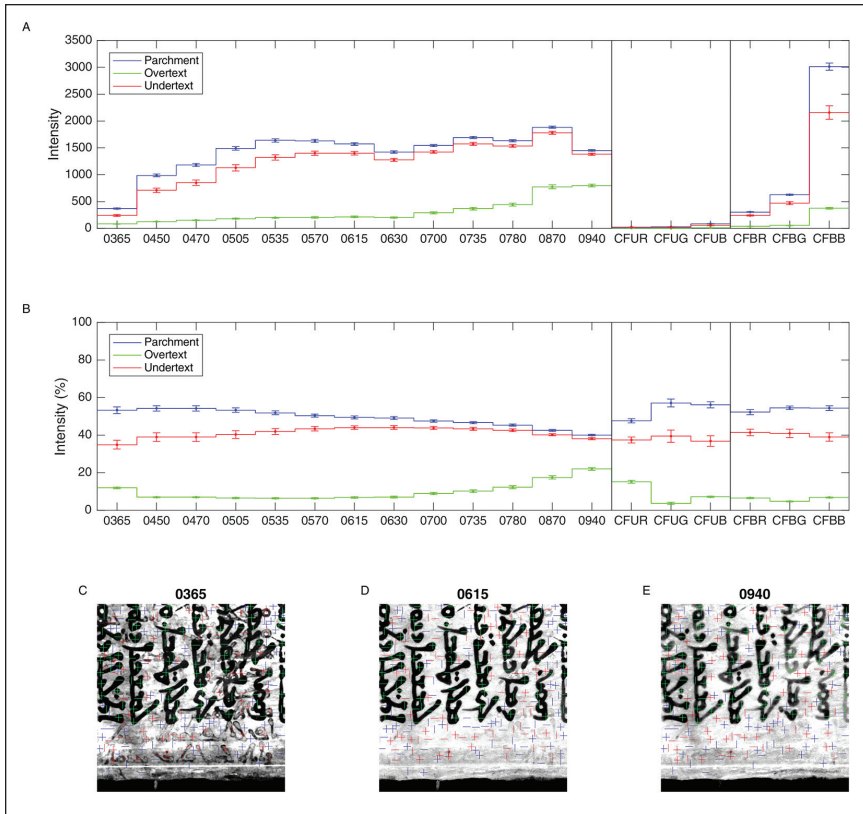


FIGURE 1. (A) Histogram showing the mean intensity values of the parchment, overtext, and undertext recorded in the 16-bit images when illuminated at specific wavelengths (365–940 nm) and using color filters (CFxy). x designates the illuminating wavelength: 365 nm U, 450 nm B. y designates the filter color (red R, green G, blue B). (B) Histogram showing the intensity proportion of the different elements. (C) 214v–221r section showing the sample points illuminated at 365 nm. (D) Illuminated at 615 nm. (E) Illuminated at 940 nm.

The nineteen images, however, clearly contain more information than can be displayed in the three channels available in a pseudocolor image, and it is likely that in something as visually complex as a palimpsest there are features in different wavelengths that might help distinguish the undertext from the parchment in different areas of the page. One approach that has become popular for dealing with relatively large numbers of images (so

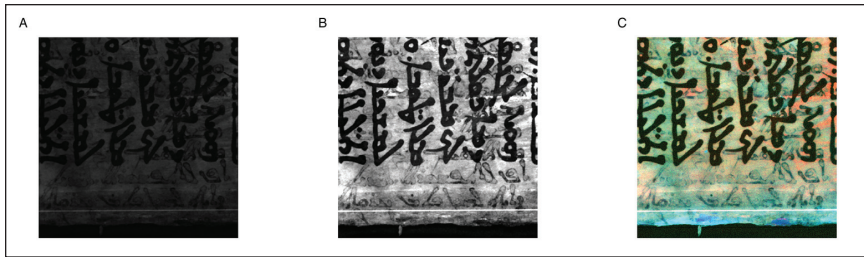


FIGURE 2. (A) 214v–221r section illuminated at 365 nm showing the full dynamic range. (B) Dynamic range set so that everything below the first percentile is set to black and everything above the ninety-ninth percentile is set to white. (C) Pseudocolor image created from 0365, CFUR, and CFBR clipped from first to ninety-ninth percentiles.

called multispectral or hyperspectral images) is dimensional reduction.²² While originally used primarily for satellite imagery, it is also being used for improving the readability of palimpsests.²³ Dimensional reduction covers a set of mathematical procedures that attempt to extract the underlying information from the individual channels such that the information alone is presented in fewer channels, for a lot of the information in the original channels is redundant and simply repeats what is found in other channels. This is clearly the case in our images since the overtext is visible in all the channels, whereas the undertext is only visible in some channels. By reducing the number of dimensions to just those that contain useful information, we should be able to increase the contrast between the different components of the image.

22 G. Shaw and D. Manolakis, “Signal Processing for Hyperspectral Image Exploitation,” *IEEE Signal Processing Magazine* 19 (2002): 12–16; doi: 10.1109/79.974715, accessed 25 May 2017.

23 Bhayro, Pormann, and Sellers, “Imaging the Syriac Galen Palimpsest”; R. L. Easton, K. T. Knox, and W. A. Christens-Barry, “Spectral Image Processing and Analysis of the Archimedes Palimpsest,” in *19th European Signal Processing Conference (EUSIPCO 2011), Proceedings*, 1440–44, <http://www.eurasip.org/Proceedings/Eusipco/Eusipco2011/papers/1569422597.pdf>, accessed 25 May 2017; F. Hollaus, M. Gau, and R. Sablatnig, “Enhancement of Multispectral Images of Degraded Documents by Employing Spatial Information,” in *12th International Conference on Document Analysis and Recognition (ICDAR 2013), Proceedings*, 145–49, doi: 10.1109/ICDAR.2013.36, accessed 25 May 2017.

There are two different classes of approaches to achieve this dimensional reduction. These differ in whether the algorithm used needs any help from a human operator in identifying the parts of the image that are to be considered interesting. Algorithms that require this initial training step are called supervised algorithms, and those that make their own judgments about what is interesting in the image are called unsupervised algorithms. In general, one might imagine that supervised algorithms should perform better (or at least more reliably), but of course they require more work from the user since the algorithm needs to be manually shown what the different parts of the image should typically look like. Unsupervised algorithms have the advantage of being completely automated, and therefore they do not require the training step and the additional user time. In addition, they can potentially find aspects of the image set of which the user was unaware. It is therefore often worth trying both approaches, with an unsupervised method being run first to help identify what is in the image and a supervised method run afterwards to try to improve the visibility of the features of interest. There are a large number of possible algorithms, and here we have chosen four examples to illustrate the major types.

The classic unsupervised dimensional reduction technique is principal component analysis (PCA).²⁴ This technique can be thought of as a linear transformation of the data into a new coordinate system such that the greatest variance of the data is along the first coordinate axis, the second greatest variation is along the second coordinate axis, and so forth. Since this variation is what can be thought of as the useful content of an image, the maximum image contrast will be obtained in a grayscale image generated from the variance along the first axis. Subsequent axes will each contain further information from the image and can be used to produce additional grayscale images. These images can be viewed in threes as pseudocolor images as described above to further enhance contrast. This process is illustrated in figure 3, and it is clear that different elements of the text are preferentially highlighted by different principal components. The values

24 K. Pearson, "On Lines and Planes of Closest Fit to Systems of Point in Space," *Philosophical Magazine* 2 (1901): 559–72.

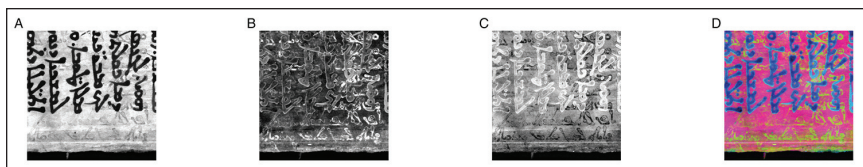


FIGURE 3. PCA-enhanced section of 214v–221r using first to ninety-ninth percentile range. (A) PC1. (B) PC2. (C) PC3. (D) Pseudocolor image created from A–C.

produced by PCA, as by most dimensional reduction techniques, are arbitrarily scaled, so they need to be mapped to the standard 0–255 intensity values permitted by most computer image formats.

There is a supervised technique that is conceptually very similar to PCA: canonical variates analysis (CVA).²⁵ This technique is also known as linear discriminant analysis (LDA). Whereas PCA takes linear combinations of the image data that maximally spreads the data within the fewest dimensions, this CVA uses pixel data that has already been split into categories by the user and produces the linear combination of the categorized data that maximally separates the categories while minimizing the spread within the categories again in a smaller number of dimensions. Once these linear combination values have been calculated, they can be applied to the rest of the image data to produce an image for which we have maximal contrast between the different categories. Once again, each dimension can be mapped to either a different grayscale image, or three such images can be combined into a pseudocolor image. This process is illustrated in figure 4, and in this case, because of the training data, it is guaranteed that the different images will separate out the specified categories (overtext, undertext, and parchment) maximally. The number of axes produced by CVA that contain useful data is always one less than the number of training categories, so only two images are produced: figure 4A has extracted the overtext, and figure 4B has extracted the undertext.

25 R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics* 7 (1936): 179–88.

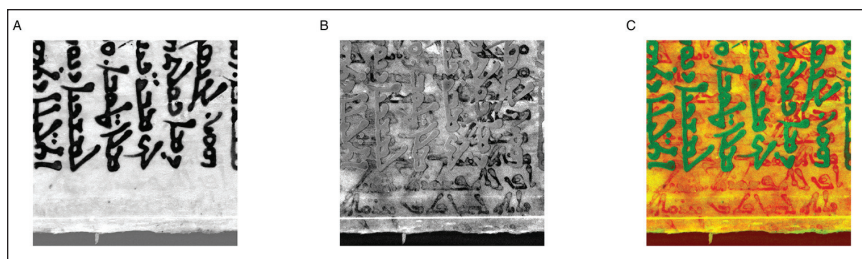


FIGURE 4. CVA/LDA-enhanced section of 214v–221r using first to ninety-ninth percentile range. (A) CV1. (B) CV2. (C) Pseudocolor image created from A–B.

PCA and CVA, and indeed many dimension reduction techniques, are linear. What that means is that the output image is a simple sum of proportions of each of the other images in a multispectral set. What varies is the method used to find what the proportions are. It is, however, perfectly possible to use nonlinear methods where this simple additive and subtractive process is not the way that the final image is generated and operations other than simple arithmetic can be used to generate the final intensity values. These nonlinear methods can improve the ability of the algorithm to highlight specific parts of the image. The downside of nonlinear approaches is that these methods often take a great deal longer to calculate than linear methods and may not actually perform as well as the linear approaches. There are supervised and unsupervised nonlinear forms, and these methods usually take additional parameters so that there is more scope for tuning their performance for a particular image enhancement task. Isomap is a good example of an unsupervised nonlinear method.²⁶ It is a distance-based dimensional reduction technique that uses the geodesic distance between locations rather than the simpler Euclidian distance that forms the basis of techniques such as PCA or multidimensional scaling (MDS).²⁷ The results of an isomap enhancement are shown in figure 5, where the isomap routine

26 J. B. Tenenbaum, V. De Silva, and J. C. Langford, “A Global Geometric Framework for Nonlinear Dimensionality Reduction,” *Science* 290 (2000): 2319.

27 W. S. Torgerson, “Multidimensional Scaling: I. Theory and Method,” *Psychometrika* 17 (1952): 401–19; doi: 10.1007/BF02288916, accessed 25 May 2017.

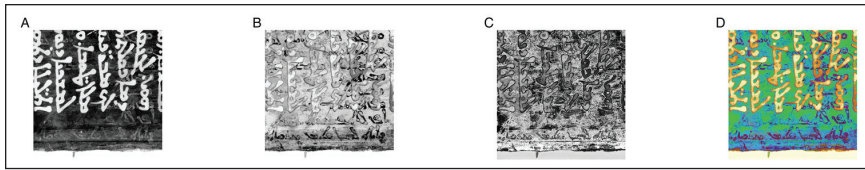


FIGURE 5. Isomap-enhanced section of 214v–221r using first to ninety-ninth percentile range. (A) Dimension 1. (B) Dimension 2. (C) Dimension 3. (D) Pseudocolor image created from A–C.

has been asked to reduce the original nineteen dimensions of the image down to three.

Generalized discriminant analysis (GDA),²⁸ also known as kernel fisher discriminant analysis, is a version of LDA that is extended such that it can use nonlinear mappings to improve the separation of the training data. This essentially uses a suitable nonlinear function to transform the data before maximizing the spread of the classes. In theory, the option of having a nonlinear function should enhance the separation, but the choice of function is important, and there is no guarantee that the performance will be better than a linear alternative. Figure 6 shows the results of using GDA with a Gaussian kernel to allow the neighborhood of a value to have an effect. This illustrates how a nonlinear method has produced a very high-contrast image; however, whether such an image is easier to read is an altogether more subjective decision, and this shows the importance of trying a number of different approaches.

As important as the choice of algorithm to increase the legibility of the manuscript pages is the choice of implementation. The images shown here were all produced using the Matlab Toolbox for Dimensionality Reduction.²⁹

28 S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Mullers, "Fisher Discriminant Analysis with Kernels," in *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop* (New York: Institute of Electrical and Electronics Engineers, 1999), 41–48, doi: 10.1109/NNSP.1999.788121, accessed 25 May 2017.

29 L. Maaten and G. Hinton, "Visualizing Data Using t-SNE," *Journal of Machine Learning Research* 9 (2008): 2579–2605; L. Van Der Maaten, E. Postma, and J. Van den Herik, "Dimensionality Reduction: A Comparative Review," *Journal of Machine Learning Research* 10 (2009): 66–71.

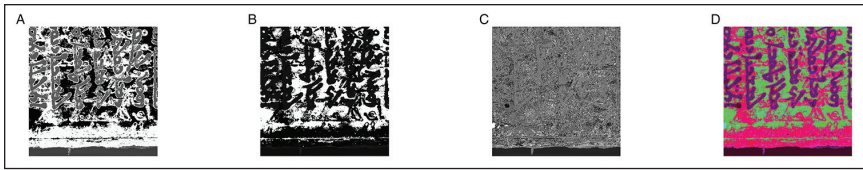


FIGURE 6. GDA-enhanced section of 214v–221r using first to ninety-ninth percentile range. (A) Dimension 1. (B) Dimension 2. (C) Dimension 3. (D) Pseudocolor image created from A–C.

Using Matlab (www.mathworks.com) is a popular option for multispectral image handling because it allows rapid implementation and, as a general-purpose numerical computing environment, many scientists are already familiar with how to use it. It relies on programming custom scripts, which provides a great deal of flexibility and makes it easy to repeat a standardized set of operations without human input. Thus, it is comparatively easy to process all two-hundred-plus leaves of the SGP over a weekend without any user intervention. However, it is not an ideal environment for scholars who wish to read the SGP. There are two problems. First, the learning curve for Matlab is relatively steep, and thus learning how to get the most out of it requires more time and expertise than is desirable. Second, it is relatively slow, particularly when being used interactively as opposed to using it to run scripts untended. There is a considerable element of personal preference in any decision about maximizing the readability of a document, so what is needed is a way to access these sophisticated methods in a rapid, interactive format that is relatively simple to learn and experiment with. To help achieve this, we are porting a range of dimension reduction techniques to a custom written piece of software called GalenQt. This is a high-speed, interactive program that allows the user to try out different options in terms of image enhancement. This enhancement encompasses both the simple options of changing the brightness, contrast, and color of images to match the user's preferences, and the more complex choice of a suitable supervised or unsupervised dimensional reduction technique. For supervised techniques, the training sets can be specified interactively by the user, and the user can mix the results from several different enhancement options in a customizable fashion. This is very much a work in progress, but figure 7 shows a screenshot of the

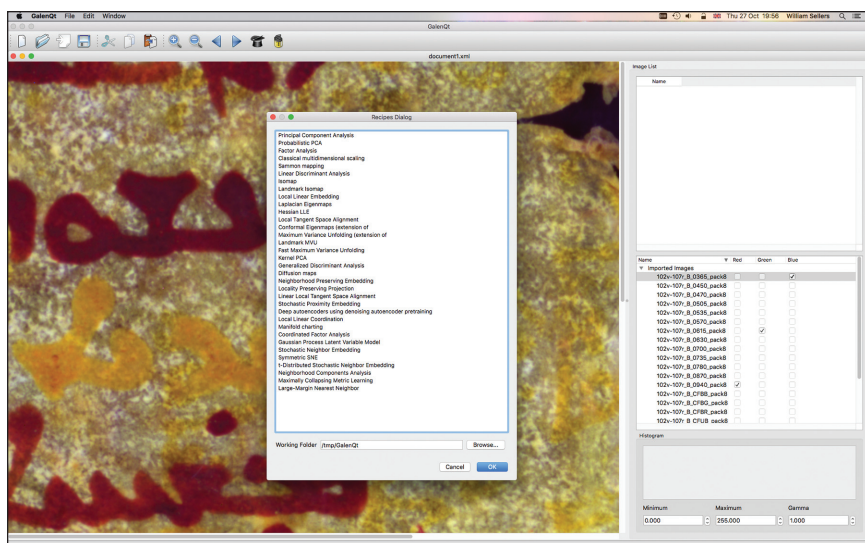


FIGURE 7. Screenshot of GalenQt, a GUI-driven multidimensional image enhancement toolbox.

current version. The program is open source, freely available, and written in C++ using the Qt library to allow it to run on all major operating systems with minimum changes in the code. We expect a public release in 2018.

Improving the readability of the undertext is an important step in deciphering it. Yet, the methodologies and tools described above need to be combined with a range of philological methods that allow us to reconstruct the text of the SGP with the help of additional evidence from MS BL and the Greek source text.

Philology

The image analysis just described considerably improved the visibility of the undertext in the SGP. In the following, we will describe the philological tools and the methods that we developed to facilitate the identification and the transcription of the undertext. To illustrate the extent of the results that can be achieved, we will present two unedited passages from the Galenic treatise in Syriac that have been completely reconstructed, mainly on the basis of the Greek text.

TOOLS

In order to study the SGP, we created a bilingual corpus of Books 6 to 8 with the SketchEngine (<http://sketchengine.co.uk>), a corpus management and query system for natural language processing.³⁰ The platform was able to handle the first corpora in Ancient Greek and Syriac using UTF-8 character encoding, despite the absence of any built-in linguistic processing for these languages. We based our bilingual corpus on the Greek version edited by Kühn (<http://www.graeco-arabic-studies.org/texts.html>) and the transcription of MS BL. After having prepared our parallel texts by manual alignment, the data files were configured and uploaded successfully onto the platform. To demonstrate the potential of utilizing SketchEngine for our research, we shall present here three of its features and practical applications.

SketchEngine enables one to search texts by query types (word, phrase, sequence of characters, etc.) and to generate simple or parallel concordances. A statistical dictionary automatically highlights translation pairs (in Greek and Syriac in our case), and thus makes it easier to survey the outputs extracted from bilingual corpora. The lists of Greek-Syriac correspondences that are created automatically do not always achieve a high accuracy, but in many cases, the dictionary performs quite well (fig. 8). SketchEngine also lets one select and combine more refined search criteria. One can, for instance, search for a word within a text or by context. The latter option is particularly relevant if one is interested in the co-occurrence of two keywords. We exploited this feature to identify a passage in the SGP (fig. 9). Another example is the collocation search, which offers the possibility to display a list of all the potential lexical combinations around a specific term to find common expressions within a corpus. The amounts of how often the collocations appear are displayed, and there is excellent cross-referencing to the concordance (fig. 10).

30 For a general description, see A. Kilgariff, P. Rychlý, P. Smrz, and D. Tugwell, "The Sketch Engine," in *Euralex 2004: Proceedings of the Eleventh EURALEX International Congress*, ed. G. Williams and S. Vessiers (Lorient: Université de Bretagne Sud, 2004), 105–16; A. Kilgariff, V. Baisa, J. Bušta, M. Jakubiček, V. Kovář, J. Michelfeit, P. Rychlý, and V. Suchomel, "The Sketch Engine: Ten Years on," *Lexicography* 1 (2014): 7–36.

FIGURE 8. Simple search for the word *σέρφον* (wormwood) in the bilingual corpus of Books 6–8; the search term appears in red on the left, and the statistical dictionary automatically highlights possible Syriac correspondences in yellow.

FIGURE 9. Simple search for the word ḥḥ (something) within the context of ḥḥḥḥḥḥ (heat) in a range of five tokens before or after.

FIGURE 10. Partial list of the collocation candidates for *maður* (he is) in a range of five tokens before or after.

SketchEngine functions well not only when exploring parallel texts, but also when browsing the content of a monolingual corpus. Therefore, we also built three searchable corpora for Sergius's Syriac introductions to Books 6 to 8, the Greek text of Books 1 to 5, and Books 9 to 11. The exploration of all these sources is thereby greatly facilitated and supports our research in a number of ways, whether for the study of the Galenic lexicon and the translation techniques or for identifying and transcribing the undertext.

The bilingual corpus for Books 6 to 8 is updated and improved progressively with the support of the technical team of SketchEngine and through exchanges with international scholars. For purposes of optimizing its use, we initiated a collaboration with Bastien Kindt of the GREgORI Project (Université catholique de Louvain), which provides automatic processing tools for languages of the Christian Orient.³¹ The long-term objective is to produce lemmatized versions for Books 6 to 8 of the *Simples* in order to be able to search for all forms of one word simultaneously in SketchEngine.

The lemmatization is being realized through Unitex, a collection of programs used for the analysis and processing of texts developed at the Laboratoire d'Automatique Documentaire et Linguistique (LADL) in Paris (<http://unitexgramlab.org/>). The operation requires one to apply electronic grammars and dictionaries, which in our case include the Dictionnaire Automatique Grec (Automated Greek Lexicon, DAG) of the GREgORI Project.³² The lemmatization interface under Unitex assigns all the possible grammatical analysis for each form of a text and represents the structure of

31 Full title: GREgORI—Softwares, linguistic data and tagged corpus for ancient GREEk and ORiental languages (directed by B. Coulie) (www.uclouvain.be/gregori-project). About the genesis of the project (formerly the Research Project in Greek Lexicology, *Projet de Recherche en Lexicologie Grecque*, PRLG), see B. Coulie, "La lemmatisation des textes grecs et byzantins: une approche particulière de la langue et des auteurs," *Byzantion* 66 (1996): 35–54.

32 About the method and the dictionary, see L. Kevers and B. Kindt, "Vers un concordanceur-lemmatisateur en ligne du grec ancien," *L'antiquité classique* 73 (2004): 203–13; B. Kindt, "La lemmatisation des sources patristiques et byzantines au service d'une description lexicale du grec ancien: Les principes de formulation des lemmes du Dictionnaire Automatique Grec (D.A.G.)," *Byzantion* 74 (2004): 213–72; B. Kindt and M. Pirard, "De Nazianze à Ninive: La couverture lexicale du Dictionnaire Automatique Grec," in *Philokappadox. In memoriam Justin Mossay*, ed. V. Somers and P. Yannopoulos, *Orientalia Lovaniensia Analecta* 25; Bibliothèque de Byzantion, 14 (Leuven: Peeters, 2016), 49–77.

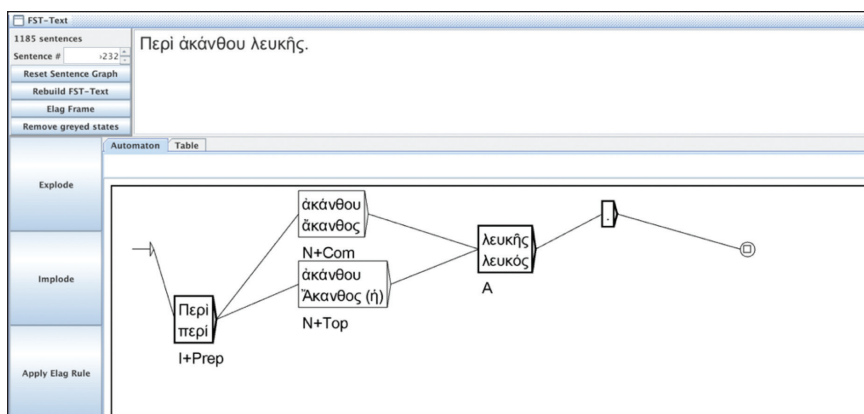


FIGURE 11. Graph example generated under Uitex.

a sentence as an editable graph (fig. 11). A disambiguation is then applicable by using automated tools or by deleting the inappropriate results manually for obtaining the best possible outcome. When processing is complete, the generated files can be exported from Uitex under various formats and converted to be acceptable for SketchEngine. At present, the lemmatization of Books 6 to 8 of the *Similes* in Greek has been achieved.³³ The results are conclusive, and queries can now be performed by lemmas on the platform (fig. 12).

In addition to the data files created for the use of SketchEngine, we have produced various working materials for Books 6 to 8 in PDF format: a lemmatized concordance, a list of corrections to Kühn's edition, a list of lemmas, a lemmas frequency index, and a parts of speech list. The same type of material is also available for the beginning of Book 1 (chapters 1–10), which was processed initially on a test basis. An interface hosted by the Centre de traitement automatique du langage (CENTAL) (<https://uclouvain.be/fr/instituts-recherche/ilc/cental>) also allows online explorations of

33 The lemmatized Greek concordance of Books 6 to 8 is now available online in free access on the GREgORI project website, <https://uclouvain.be/fr/instituts-recherche/incal/ciol/the-concordances-of-the-gregori-project.html>, accessed 7 February 2018.

τὸ κρινόμενον φάρμακον **λαμβάνοι**, τὸ κατὰ τὴν
 ἐκ τῶν μετάλλων **λαμβάνεται**, κάλλιον ἔδοξε μοι
 αὐτοῖ μηδεπώποτε θεασάμενοι. **λαμβάνουσι** μὲν γὰρ τὰ
 ἢ τρίψις αὐτοῦ **εἰληφυῖα**, ὥσπερ τὸ σαρκωτικόν
 τε τοῦ σώματος **λαμβάνον** ὀλκῇ δραχμῆς μιᾶς
 φυλακτέον ἐν ἐδέσματι **λαβεῖν** ἢ ποτῶ. πρὸς
 πείρας τε μόνης **λαμβάνεται** καὶ παντάπασιν ἀμέθοδος
 τὰ φύλλα ξηρὰ **λαβῶν**, ἢ τὴν ρίζαν
 ὀρώντι, κᾶπειδ' ὀλίγον **ληφθὲν** ἐπιπλεῖστον ἐκτείνηται τοῦ
 πεπερέως ἡδονῆς ἔνεκα **λαμβανόμενον**. ἐπισπάται δὲ καὶ

FIGURE 12. Search by lemma of λαμβάνω (to take) in Book 6 with Sketch Engine (ten results).

the entirely lemmatized corpus (<http://cental.uclouvain.be/galen/>).³⁴ As the tools generated within the framework of the collaboration between the GREgORI and the SGP projects are open source, anyone can access them.

For continuous operational improvement, we foresee the possibility of applying additional computational tools to the Syriac Galenic text (Books 6 to 8). To do so, the Manchester Project is collaborating with the ERC-funded project “Hunaynnet: Transmission of Classical Scientific and Philosophical Literature from Greek into Syriac and Arabic” (Austrian Academy of Sciences) and other colleagues across the Channel and the Atlantic who are interested in Syriac computer-aided corpus linguistics.

Methodology

The use of computer-assisted approaches helped us to explore the Galenic text quickly. The tools we developed thus simplify the identification and transcription of the undertext.

³⁴ An updated version of the interface of CENTRAL is currently in preparation and will be fully operational in the near future.

We first established which parts of the SGP contain Books 6 to 8 by using the bilingual MS BL corpus, which (as described) includes: (1) the Greek source and Syriac target text (except for some lacunae in Book 8);³⁵ (2) Sergius's Syriac introductions to Books 6, 7, and 8; and (3) Syriac lists of plants in alphabetical order. We collated the SGP undertext with the Syriac parallel text in MS BL, wherever available, and obviously this makes the reading of the palimpsest considerably easier.

After the initial phase, we determined that approximately one hundred folios of the undertext contained text that does not appear in MS BL. We then endeavored to determine whether these leaves contained text from Books 1 to 5 and Books 9 to 11. Our research has revealed that with the exception of three folios that remain completely illegible (fols. 031r/v–038v/r, 143r/v–146v/r, 191r/v–194v/r), the palimpsest contains parts of Books 2 to 9, whereas Books 1, 10, and 11 are not represented.³⁶ We were able to establish this by comparing the Syriac in the SGP with the Greek source text, available in our electronic corpus. In rare cases when the content of the palimpsest has no equivalent, we referred to the Greek manuscript tradition. There is, for instance, a passage in the SGP, folios 184v–185r in Book 9, that mentions the name of an Armenian city and contains a description of the letter *digamma*; this passage does not appear in Kühn's edition, but is present in one Greek manuscript—namely, Rome, Vatican Library, Urbinas Gr. 67.³⁷

We have already published some preliminary results elsewhere.³⁸ It is, however, useful to illustrate our methodology by looking at two further concrete examples, one from Book 5, for which we do not have the Syriac parallel text in MS BL, and the other from Book 8. In both cases, we

35 See further Example 2.

36 On the contents of the undertext, see Afif et al., “The Syriac Galen Palimpsest: A Tale of Two Texts,” in this volume.

37 M. Martelli, “Galeno grammatico sui nomi stranieri e il digamma: un passo inedito dal IX libro del trattato Sui medicamenti semplici,” *A.I.O.N.: Annali dell'Università degli Studi di Napoli “L'Orientale,” Dipartimento di Studi del Mondo Classico e del Mediterraneo Antico, Sezione filologico-letteraria* 34 (2012): 131–46.

38 Afif et al., “Continuing Research on the Syriac Galen Palimpsest”; N. Afif, S. Bhayro, P. E. Pormann, W. I. Sellers, and N. Smelova, “The Syriac Galen Palimpsest: Research Methods and Latest Discoveries,” *Comparative Oriental Manuscript Studies Bulletin* 2 (2016): 5–16.

reconstructed the undertext by considering paleographic evidence (the shape of a character), the manuscript layout (blank spaces and the spaces that supposedly contain the titles that may not be visible) and the proposed number of missing letters. Instead of a diplomatic transcription, we provide a reconstructed critical edition of the Syriac texts. The conventions of edition are voluntarily simplified: reconstructed terms as well as uncertain and erased letters or words are indicated in square brackets.

Let us begin with the first example, where Galen briefly discusses the four elements and elemental qualities, and how to name different parts of the body.

EXAMPLE 1: SGP FOLIOS 175V–178R

Column B, lines 2–21, covers the beginning of Book 5 (ed. Kühn XI.704–5), which is not represented in MS BL. As column A corresponds to the end of the *pinax* (list of chapters), the unreadable section in column B, lines 1–2, may include a rubricated title written in an ink different from that of the main text. The following lines are mostly decipherable despite the covering overtext and are reconstructed according to the Greek text.

ܠܬܠܡܕ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ [1] [. . .]
 ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ [2] ܕܥܠܡ
 ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ [3] ܕܥܠܡ
 ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ [4] ܕܥܠܡ
 ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ [5] ܕܥܠܡ
 ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ [6] ܕܥܠܡ
 ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ [7] ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ

39 The verb ܥܠܡ (compose) in this context is a possible reading, but the meaning differs from the Greek ἐνστησάμενος (to place in or to begin).

40 The word order in Syriac differs from the Greek: ܕܥܠܡ ܕܥܠܡ ܕܥܠܡ (warm, cold, and dry) for καὶ ξηρὸν καὶ θερμὸν καὶ ψυχρὸν (dry, warm, and cold). See also line 21: ܕܥܠܡ ܕܥܠܡ (warm and cold) for καὶ ψυχρὰ καὶ θερμὰ (cold and warm).

41 In place of ܕܥܠܡ the scribe wrote ܕܥܠܡ but he corrected the ܥ to ܕ under the write line.

[1] This is the fifth book that I have composed on the powers of simple drugs. [2] First, I shall mention the things which I have previously shown are necessary even now; [3] I shall begin with the elements, [4] namely water, fire, earth, and air, [5] that others are calling according to their qualities,⁴³ wet, warm, cold, and dry, [6] for which the qualities are therefore moistness, dryness, warmth, and coldness. [7] The bodies [of these qualities] that are common for all things called elements, [8] are named moist, dry, warm, or cold, according to which one of the [qualities] prevails in them.

[1] Πέμπτον τοῦτον ἐνστησάμενος λόγον ὑπὲρ τῆς τῶν ἀπλῶν
φαρμάκων δυνάμεως, [2] ἀναμνήσω πρῶτον ὧν ἐν τοῖς ἔμπροσθεν
ἀπέδειξα χρησίμων εἰς τὰ παρόντα, [3] ποιησάμενος τὴν ἀρχὴν
ἀπὸ τῶν στοιχείων, [4] ὕδατός τε καὶ πυρὸς, ἀέρος τε καὶ γῆς,
[5] ἅπερ ἔνιοι παρονομάζοντες ἀπὸ τῶν ποιοτήτων ὕγρον καὶ
ξηρὸν καὶ θερμὸν καὶ ψυχρὸν εἶναί φασιν. [6] αὗται μὲν οὖν αἱ
ποιότητες, ὕγρότης τε καὶ ξηρότης, θερμότης τε καὶ ψυχρότης εἰσί.
[7] τὰ δ' ἀπ' αὐτῶν παρονομαζόμενα σώματα τὰ τε κοινὰ πάντων
ἐστὶ στοιχεῖα καὶ τὰ κατ' ἐπικράτησιν τινος τούτων, [8] ὕγρὰ καὶ
ξηρὰ καὶ ψυχρὰ καὶ θερμὰ προσαγορευόμενα.

EXAMPLE 2: SGP FOLIOS 166R–171V

The second example comes from Book 8, in which Galen describes different types of mullein, called *phlómos* in Greek—namely, the white, black, and

42 The blurred initial letter is read as a ⲟ for ⲟⲩⲟⲩ followed by ⲕⲁⲩ ⲙⲁ ⲕⲁⲩ (according to the part of preeminence) translating $\kappa\alpha\iota \tau\acute{\alpha} \kappa\alpha\tau' \epsilon\pi\iota\kappa\rho\acute{\alpha}\tau\eta\sigma\acute{\iota}\nu$ (and according to predominance).

43 The term *κρίσις* refers to ποιότης (quality) in Galen's terminology.

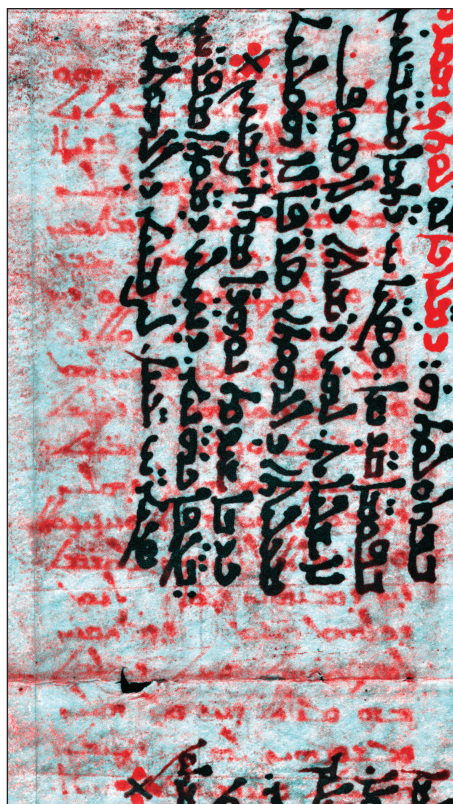


FIGURE 13. A pseudocolor image of SGP folio 175v–178r, column B, lines 2–21.

wild varieties, and that called *thryallis*.⁴⁴ Here our conjectures are based on the results provided by searching recurrent Greco-Syriac lexical pairs in the bilingual corpus. Instead of a diplomatic transcription, we provide a reconstructed critical edition of the Syriac texts. We have added diacritical marks (notably the plural marker *seyame*) and punctuation where required. Text in square brackets cannot be read and is reconstructed.

⁴⁴ *Phlomos* (ἡ φλόμος) refers to various species of mullein. The plant is described by Dioscorides (*Materia Medica* 4.103), who distinguishes four types. In the *Simples*, mullein is also compared to ἄρκτιον (bearwort) (Book 6, 1/58); see M. Ullmann, *Wörterbuch zu den griechisch-arabischen Übersetzungen des 9. Jahrhunderts* (Wiesbaden: Harrassowitz, 2002), 736.

[illegible]

⁵³ The reading حمى is supported by a passage in Book 6 mentioning *phlómos*: חמלם וסחב סכא דהויה רבה למי (MS BL, fol. 18r, ll. 32–33).

54. אִתְּךָ מִתְּחִלָּה חָזַר אֶל-לִבִּי וְהִלַּחֲמֵנוּ. וְכֵן חָזַר לִשְׁכַּתִּי כִּשְׁמִי.
 [12] אֲנִי עָמֵד בְּחִצְתִּי וְחָסָה אֵלַי כִּי אֶלֶף חֲסִידֶיךָ [13] בְּיָמֵי-עֲדֻמִּי וְהָיָה
 שִׁלְשֵׁל מִבֶּן חֲמֵשׁ עָשָׂר [14] (וְחִסְדְּךָ בְּ-אֶרֶץ) 55 בְּיָמֵי הַגָּלוּת וְכִשְׁמִי חָסָה
 בְּיָמֵי הַגָּלוּת. וְכֵן חָזַר לִשְׁכַּתִּי וְהִלַּחֲמֵנוּ. [15] וְהָיָה עָמֵד בְּיָמֵי
 חֲמִשָּׁה אֶלֶף חֲסִידֶיךָ. [16] וְכֵן חָזַר לִשְׁכַּתִּי וְהִלַּחֲמֵנוּ שִׁלְשֵׁל מִבֶּן חֲמֵשׁ עָשָׂר
 וְכֵן חָזַר לִשְׁכַּתִּי.

[1] [On mullein.] This plant also has many species. [2] One of them has white leaves [3] and other [species] has black [leaves]. [4] Regarding the one with white leaves, [5] a male [plant] has narrow leaves and [6] a female [plant] has wide and large [leaves]. [7] Besides them, there is also another one called wild mullein, [8] whose leaves are yellow and pure [like gold]. [9] And apart from these, there is yet another one [10] that is specifically called *phlomis* or *thryallis*, that is verbasicum. [11] The root of the two former tastes astringent. Therefore, it is beneficial for diseases accompanied by flowing [*ḥašṣe rdūyē*]. [12] People also gargle with it for [the relief of] the toothaches. [13] Their leaves have a diaphoretic power. [14] Similar [action have] the leaves of the other [plants] of the same kind, especially the one with golden leaves. [15] Leaves of all of them [can] also dye the hair red, [16] because all these leaves have a drying power which is moderately thinning.

[1] Φλόμος, ἔνιοι δὲ διὰ τοῦ π γράφουσι τε καὶ λέγουσιν. [2] ἡ μὲν τίς ἐστι τοῖς φύλλοις λευκὴ, [3] ἡ δὲ μέλαινα. [4] καὶ τῆς λευκῆς ἡ μὲν ἄρρην, [5] ἥ τὰ φύλλα στενώτερα, [6] ἡ δὲ θῆλυς, ἥ μείζω. [7] καὶ πρὸς ταύταις ἑτέρα φλόμος ἀγρία προσαγορευομένη, [8] χρυσοειδὴς δ' ἐστὶ καὶ μηλίζουσα τοῖς ἄνθεσιν. [9] ἄλλη δέ τις ἐπὶ ταύταις, [10] ἣν ἰδίως φλομίδα καὶ θρυαλλίδα προσαγορεύουσι.

54 𐤒𐤓𐤓] is our conjecture; SGP clearly has 𐤒𐤓𐤓. Sergius translates “*τῶν ῥοδῶδων παθῶν*” as “𐤒𐤓𐤓 𐤒𐤓𐤓” (Book 8, chapter 11; Kühn XII.90; MS BL, fol. 58v, ll. 17–18) and “*τοῖς ῥοδῶδεσι πάθεισιν*” as “𐤒𐤓𐤓 𐤒𐤓𐤓” (Book 7, chapter 22; Kühn XII.24; MS BL, fol. 40r, l. 7).

55 The text is damaged, but οὐτὼ δὲ καὶ is usually translated by 𐤀𐤓 𐤁𐤏 𐤓𐤕𐤍 in Books 6 to 8 and could fill the gap adequately.

[11] τῶν μὲν οὖν πρώτων δυοῖν ἡ ρίζα στρυφνὴ γενομένη οὔσα τοῖς ῥοώδεσι πάθεσιν ἀρήγει. [12] διακλύζονται δ' αὐτὴν ἔνιοι καὶ πρὸς ὀδόντων ἀλγήματα. [13] τὰ μέντοι φύλλα διαφορητικῆς ἐστὶ δυνάμεως. [14] οὕτω δὲ καὶ τὰ τῶν ἄλλων φύλλα καὶ μάλιστα τῆς χρυσιζούσης τοῖς ἄνθεσιν, [15] ἥ καὶ τὰς τρίχας ξανθίζουσιν. [16] ἀπασῶν γὰρ τὰ φύλλα ξηραντικῆς τε καὶ μετρίως ῥυπτικῆς ἐστὶ δυνάμεως.

These two examples demonstrate the power of our philological methods. In the case of the first example, we did not have the Syriac text preserved in MS BL. Therefore, only through our computational tools were we able to reconstruct the Syriac text. But even in the second example, there is a lacuna in MS BL that can be filled with the help of the SGP. Again, this is only possible through the parallel corpus in SketchEngine that allowed us to find Syriac equivalents to the Greek text that we have identified. As we decipher more text from outside Books 6 to 8 of the Syriac version of Galen's *On Simple Drugs*, we shall enlarge the textual basis of our corpus and thus refine this powerful tool. Yet further developments of our methodology still lie in the future, and it is to these possibilities that we now turn in our conclusions.

Conclusion

The Manchester project started in September 2015, nearly three years ago. Since then, we have made significant progress. We notably have built provisional tools for image analysis and philological study. Yet we also refined our codicological methodology, as our previous article showed. All this allowed us to identify the majority of the leaves in the SGP, including many of those that are extremely hard to read even after significant efforts at image enhancement. We are pleased by our progress, but the main task still lies ahead: for Books 6 to 8, to collate the SGP with MS BL and thus produce a critical text; and for the remaining books to edit the SGP and provide a

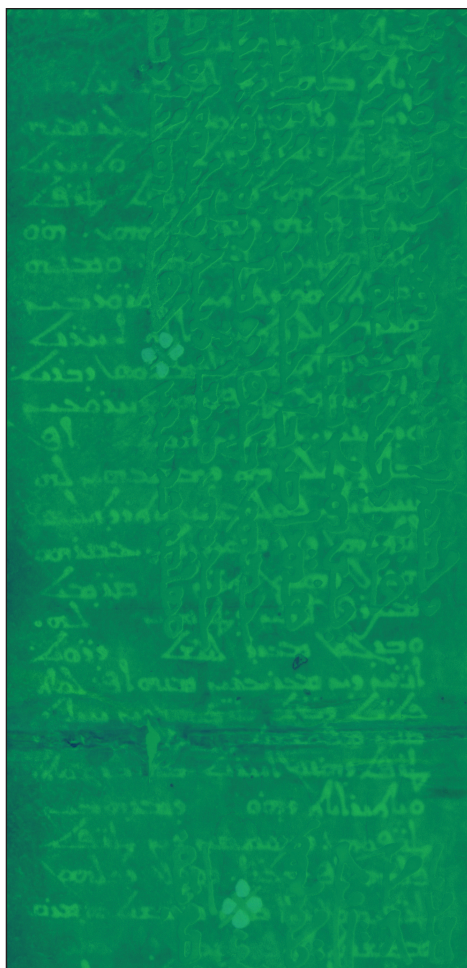


FIGURE 14. SGP folios 166r-171v, column B, lines 2-28.

first edition of the Syriac translation that it contains. It is to these two tasks (in this order) that our attention now turns.

In this article, we have described our methodologies and the tools that we have created. In the area of image analysis, the combination of unsupervised and supervised methods has allowed us to make great strides in deciphering the SGP. Moreover, we expect that our approach will also be useful

in other contexts where scholars endeavor to read erased texts, especially in the area of palimpsests. In particular, we shall share our GalenQt tool freely in the near future and look forward to receiving feedback from colleagues from different disciplines, and we are excited to see how others will further enhance these open-source tools to adapt them to their own needs.

Similarly, the SketchEngine tool we created is already quite powerful. The lemmatization of the Greek now allows us to search for a verb or a noun in its various forms, and thus create more accurate concordances or find translation pairs with greater assurance. SketchEngine, to be sure, is proprietary, yet we shall make all our raw data, including the aligned corpora, available freely under a Creative Commons license. In fact, we are very grateful to the team at SketchEngine, who helped us tremendously getting their tool to work with classical Greek and Syriac, two languages that have hitherto not been present on their platform and offered their own challenges.

There are, of course, many challenges on the horizon. Once we have transcribed more text from Books 2 to 5, we hope to breach the question of whether the SGP's original manuscript contained the same translation in the first part (Books 1 to 5) as it did in the second (Books 6 to 11). To recap, Books 6 to 8 in MS BL comprise Sergius of Rēš 'Aynā's translation, but Ḥunayn only recorded that Sergius rendered the second part into Syriac, not the first. Yet, Bhayro and Brock argued that it is likely Sergius also translated the first part. The question then is: does the SGP contain this translation? To answer this question, we will need to compare the translation in the first part with that of the second, notably by analyzing translation lexicon and syntax. Here our growing bilingual corpus will be of tremendous usefulness.

We also hope to make further progress in image capture and analysis. The present set of data, even when thoroughly analyzed, still contains many leaves that, although identified, cannot be read for the most part. This is especially problematic for parts for which we do not have MS BL and cannot use this parallel source to reconstruct the text. Here we shall pursue a two-pronged strategy. On the one hand, we hope to enhance the quality of our processed images through more detailed analysis with our present methods, for instance by focusing on smaller segments of the page. On the other hand, we hope to be able to image the SGP with novel techniques, namely through X-ray fluorescence (XRF) and computed tomography (CT). XRF would allow

us to see the ink of the undertext much more clearly, albeit indistinguishable at first from the overtext, provided that it has a good iron content (which appears likely). CT in a synchrotron would possibly allow us to construct a three-dimensional image of the whole palimpsest, and then separate the layers of over- and undertext computationally. The latter approach is more speculative, but we are currently setting up a trial.

There are also exciting new initiatives afoot in the area of analyzing Greco-Syriac translation technique. We see our own project as part of an effort to create open-source corpora of Syriac texts, especially texts translated from Greek. Other initiatives include Syriaca.org: The Syriac Reference Portal, which aims at creating a digital library of Syriac texts, and Grigory Kessel's ERC-funded project on the "Transmission of Classical Scientific and Philosophical Literature from Greek into Syriac and Arabic," which also aims at creating multilingual corpora of translated scientific texts. To extend Syriac corpora is important, but so is to create tools to analyze them. Kessel and Pormann recently organized a workshop on Syriac lemmatization. Although there have been attempts to create a tool to lemmatize Syriac in the past, they have not yet come to fruition. We hope that through the joint efforts of colleagues from Europe and the Americas, we will make progress in this field.

In conclusion, then, this article, and the whole special issue more generally, shows the tremendous progress that has been made in the study of the Syriac Galen Palimpsest. The new image analysis and philological tools have greatly helped in determining what the undertext of the SGP contains. As always, however, current results can only serve to stimulate us to future research, as many more questions remain unanswered. The progress to date was made possible only through the collective efforts of colleagues from different disciplines and institutions. It is extremely gratifying to see such selfless collaboration where unpublished material is freely shared in a sense of scholarly "humanity" (or perhaps we should say *humanitas*) that one does not always find in today's academia, where competition for posts and resources is often fierce. But together, we are more than the sum of our parts, and together, we achieve more than any one individual could ever have done. Therefore, we vow to continue our work in this spirit of collaboration to unravel the SGP.